Testing lots of ML models • Lab 10: Try the same (a set of many models and hyperparameters) on a dataset of your choice. Explain what you tried and why, and what worked ~2 pages

What I am looking for: • Dataset: What is it? Target feature for prediction? • Is cleanup required? What choices were made? • Any obvious initial trends? with or without PCA? • Any obvious clusters that inform your model decisions?

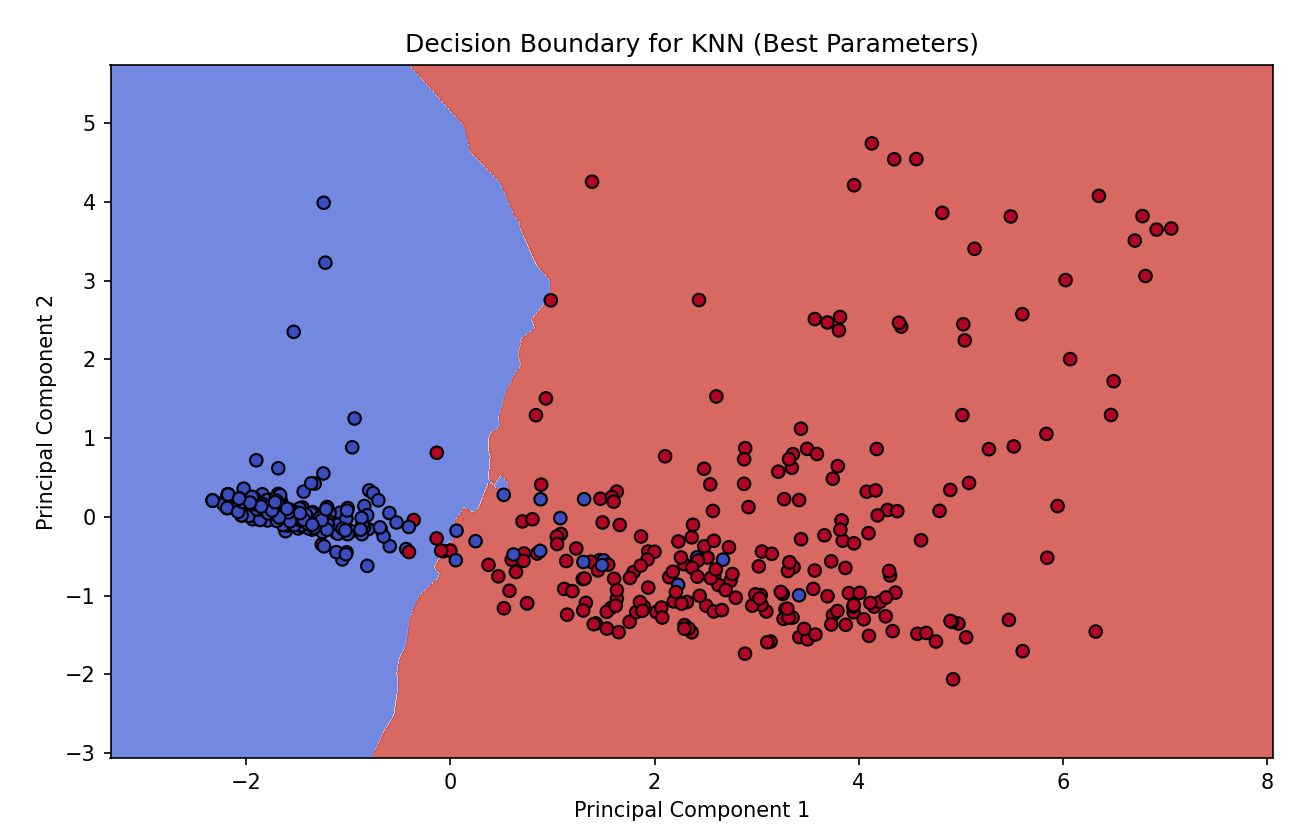
What I am looking for: • Which models tested? (~5 different\* ones) • What values of hyperparameters optimised? (gridsearch) • Compare optimised fitted model accuracy/R2 • train data to confirm not underfitting (low on train) • Test data to confirm not overfitting (high on train, low on test) • If classification, compare accuracy to F1 score

• Lab 10: Try the same (a set of many models and hyperparameters) on a dataset of your choice. Explain what you tried and why, and what worked ~2 pages

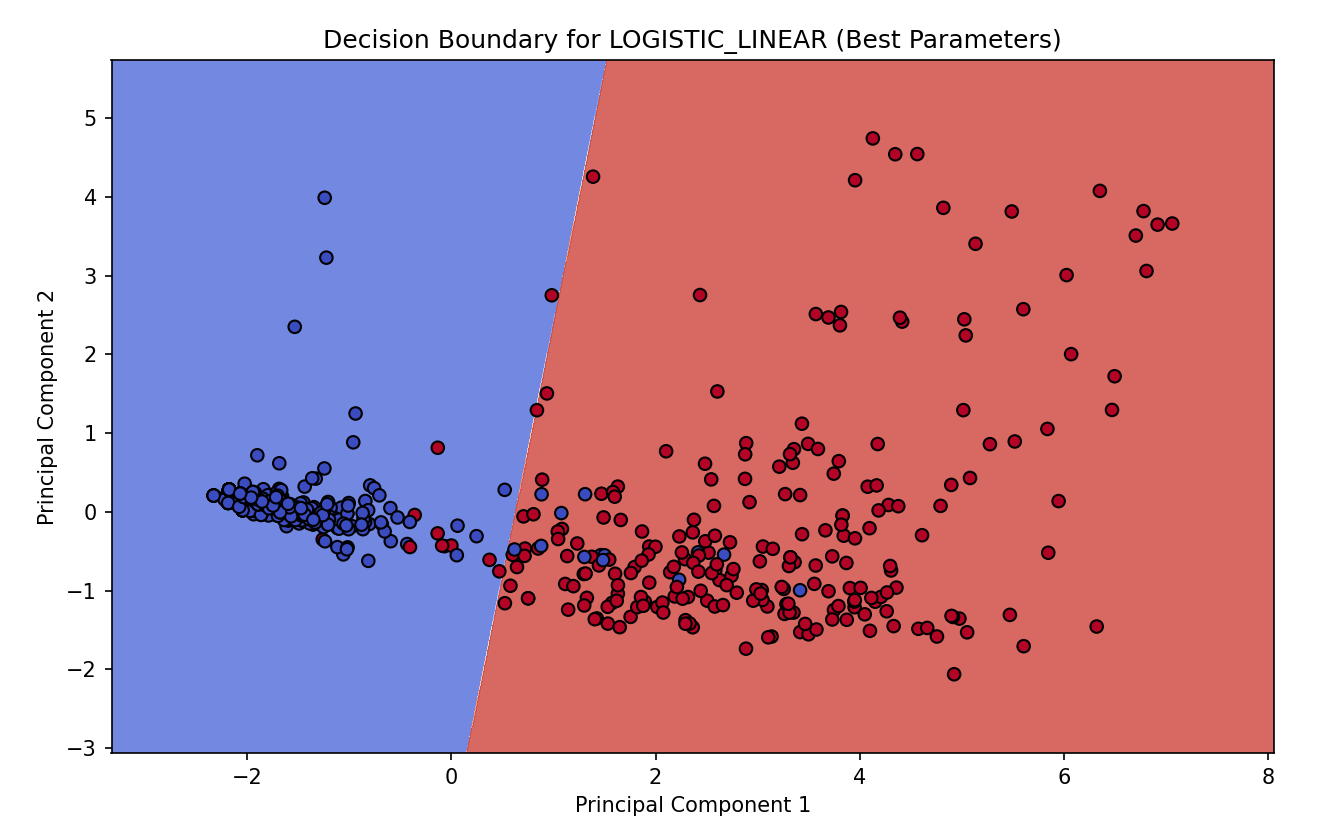
Lab 10 Report

Name: Mihir Hurwanth

Email: [bt21442@qmul.ac.uk](mailto:bt21442@qmul.ac.uk)

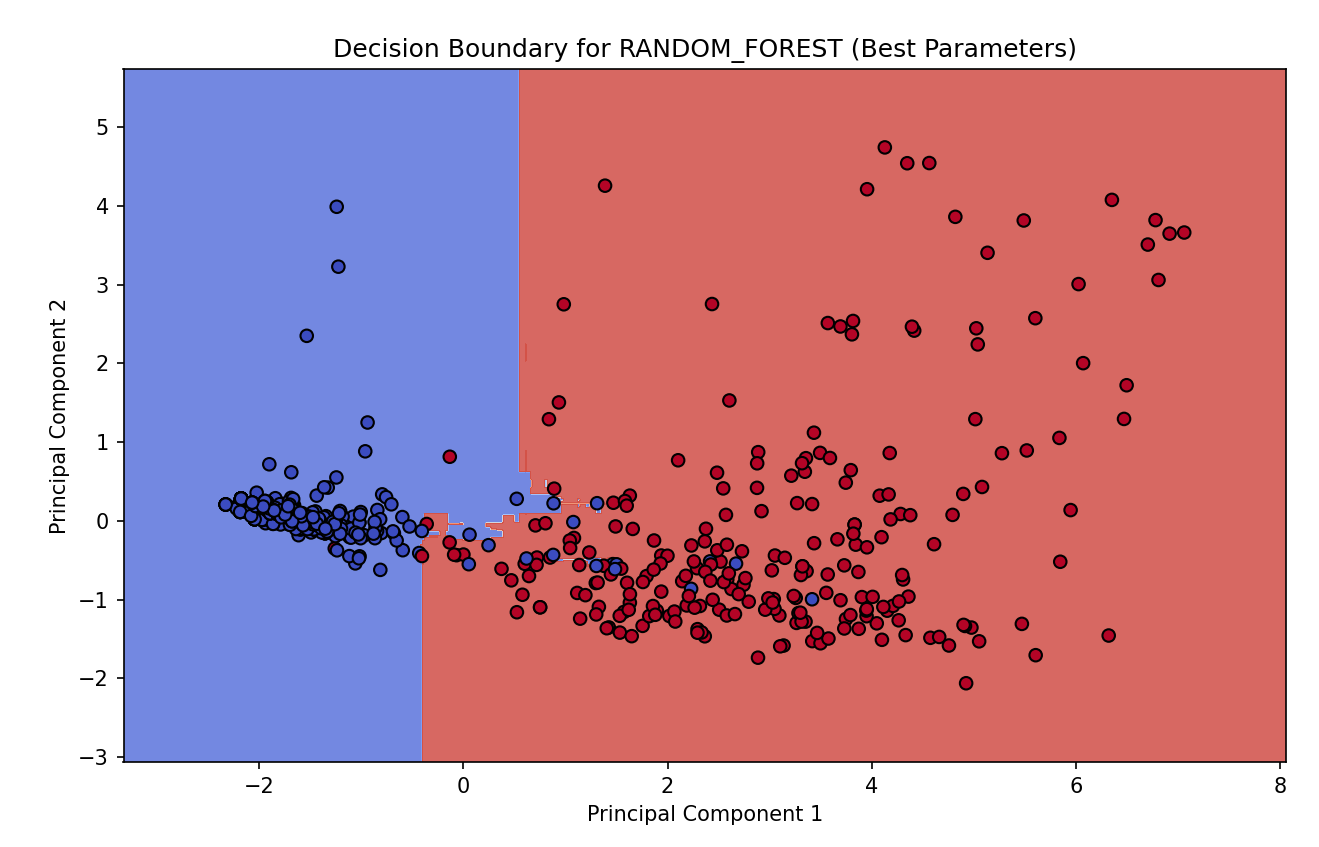
  
Parameters for KNN: {'n\_neighbors': 9, 'weights': 'uniform'}

Best Cross-Validated Accuracy for KNN: 0.9692990239574091



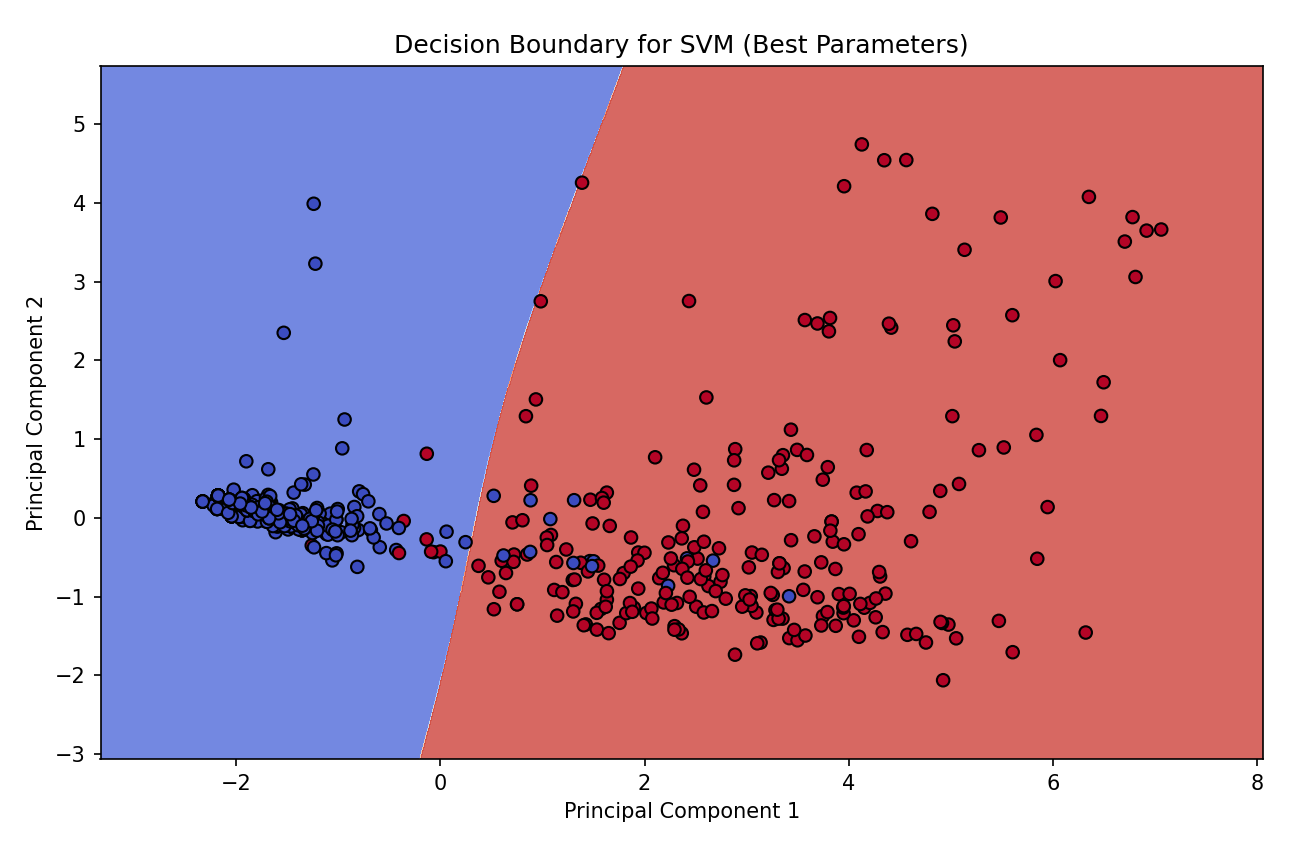
Best Parameters for LOGISTIC\_LINEAR: {'C': 0.1, 'penalty': 'l2', 'solver': 'lbfgs'}

Best Cross-Validated Accuracy for LOGISTIC\_LINEAR: 0.9672286305826677



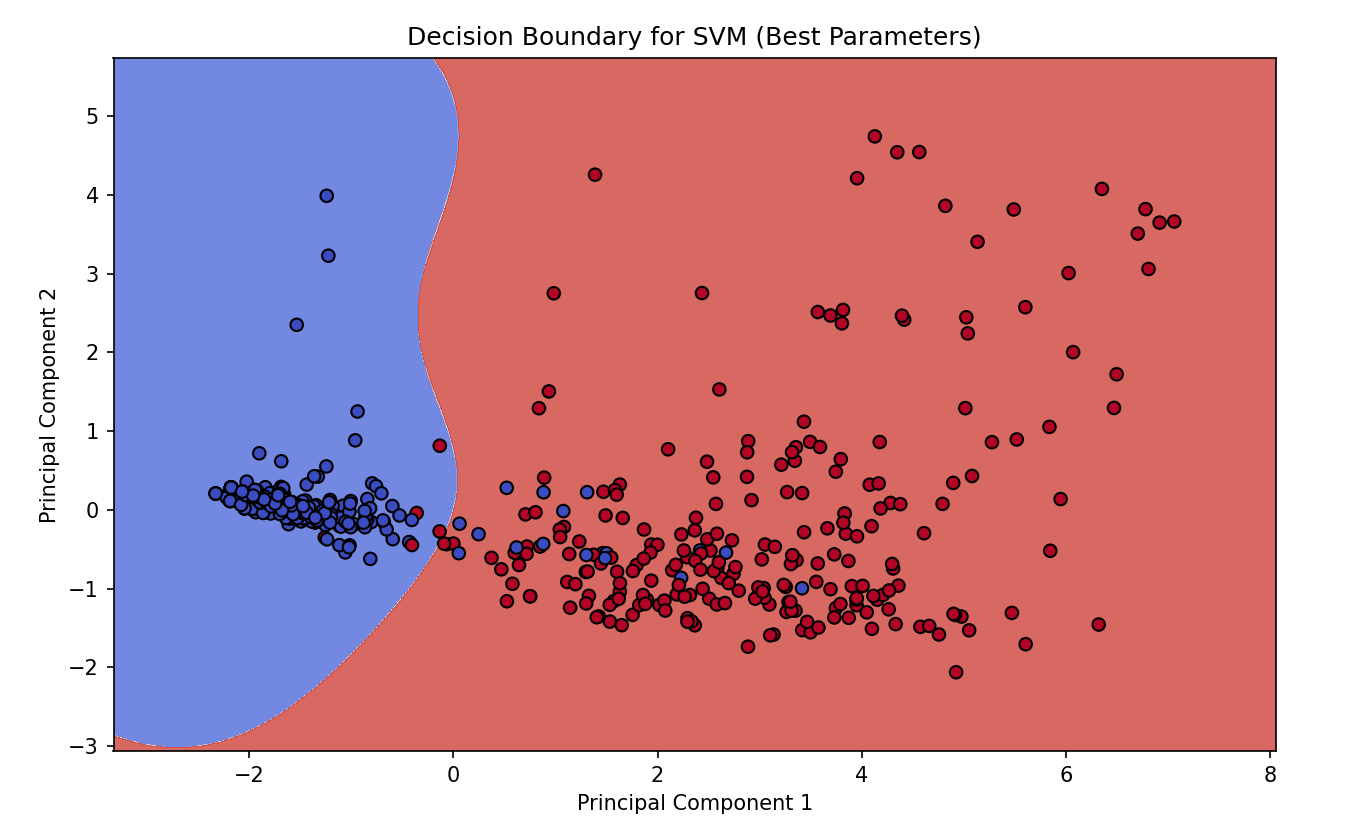
Best Parameters for RANDOM\_FOREST: {'max\_depth': 10, 'min\_samples\_split': 5, 'n\_estimators': 100}

Best Cross-Validated Accuracy for RANDOM\_FOREST: 0.9795622596864834



Best Parameters for SVM: {'C': 0.1, 'kernel': 'sigmoid'}

Best Cross-Validated Accuracy for SVM: 0.9733806566104704



Best Parameters for SVM: {'C': 10, 'kernel': 'rbf'}

Best Cross-Validated Accuracy for SVM: 0.9652173913043479